PARSING AS DEDUCTION, GRAMMAR AS LOGIC

Glyn Morrill

Departament de Llenguatges i Sistemes Informàtics
Universitat Politècnica de Barcelona

Abstract

A large part of computational linguistics is about grammar formalisms: low level formalisms such as CF-PSG, DCG and PATR-II, and higher level formalisms such as LFG, GPSG and HPSG. We seek to
- develop grammar formalisms;
- express grammars in those formalisms;
- find algorithms for NLP (parsing, generation, ...) on the basis of those formalisms;

Around ten years ago, attention was drawn to the fact that computing whether an expression is grammatical is a process of inference (on the basis of a grammar): this perspective went by the slogan 'Parsing as Deduction' (Pereira and Warren 1983). The point I want to pursue here is that if parsing can be seen as deduction, grammar can be seen as logic.

Why should we be interested in trying to see things this way? Well, with respect to the programming formalisms of computer science for example, the logical perspective has provided a methodology for design of new formalisms with remarkable properties of e.g. program error-detection (type-checking), program verification (checking that a program meets a specification), and even program synthesis (construction of a program from a specification). In AI also it has been found advantagous to develop representation formalisms on the basis of logical foundations: by keeping to a regular mathematical structure, it is possible to prove powerful results and develop powerful algorithms. It seems natural to assume that grammar formalisms are no exception to the advantages of logical treatment.

Categorial grammar as construed by Lambek (1958) provides a direct instantiation of 'Grammar as Logic': formulas are generated by directional implication connectives / and \, model theories can be given (e.g. interpretation of formulas as sets of strings), proof theories can be given (e.g. Gentzen-style sequent calculus), and soundness and completeness results

can be obtained. A grammar is simply a categorisation of words, i.e. a lexicon, and parsing is deduction of the categorisation of strings according to the logic. The elegance of this system has attracted much interest from the side of logic, but rather less from the side of linguistics, where its very mathematical simplicity suggests that it is unsuitable for the complexities of natural language grammar. And indeed this is true of the *basic* system.

Nevertheless, the methodology is preserved and applicability gained by enriching the system used for linguistic classification. This is done by adding new connectives, some familiar from logic, others specifically customised for linguistic purposes, but all being interpreted and having a logic. Each kind will be illustrated in this paper where we look at two phenomena that seem to involve both feature percolation and complex transfer of information from syntactic categories to semantic translations.

The first of these concerns semantically potent features such as gender in English, and tense, which are relevant to grammaticality (e.g. *'John walked tomorrow'; *'John walks yesterday') but also relevant semantically, e.g. tense features, encoded as verb inflection, take semantic scope at the sentence level. How is this transfer to be managed?

The second concerns pied-piping, the phenomenon whereby a relative pronoun may be accompanied by additional material from its extraction site (e.g. 'a man whom$_i$ John voted for e_i and 'a man [for whom]$_j$ John voted e_j). This puzzling construction is notoriously difficult for categorial grammar without feature percolation (Pollard 1988). We shall present a solution within the purist categorial tradition.

References

Lambek, J., 1958: 'The Mathematics of Sentence Structure', *American. Mathematical. Monthly* 65, 154-170.

Pereira, F. and D. Warren, 1983: 'Parsing as Deduction', in *Proceedings ACL*, MIT, Cambridge, Mass.

Pollard, Carl J., 1988: 'Categorial Grammar and Phrase Structure Grammar: an Excursion on the Syntax-Semantics Frontier'. In Richard T. Oehrle, Emmon Bach and Deirdre Wheeler (eds) *Categorial Grammars and Natural Language Structures*.